

ABSTRACT OF THE DISCLOSURE

A proxy partition cache (PPC) architecture and a technique for address-partitioning a proxy cache consisting of a grouping of discrete, cooperating caches (servers) is provided. Client requests for objects (files) of a given size are redirected or reassigned to a single cache in the grouping, notwithstanding the cache to which the request is made by the load-balancing mechanism (such as a Layer 4 switch) based upon load-balancing considerations. The file is then returned to the switch via the switch-designated cache for vending to the requesting client. The redirection/reassignment occurs according to a function within the cache to which the request is directed so that the switch remains freed from additional tasks that can compromise speed.

112024-0062C1